Model Selection with Graphical Representation

Graphical Analysis for Rasch and IRT models

March 15, 2021

SB Galvin

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh
School of Applied Psychology

Education:
BA Applied Psychology, PhD Candidate School of Applied Psychology, Government of Ireland Postgraduate Scholar

Research Interests:
Psychometrics, Rasch Measurement, Data Analysis, Cognitive Assessment, Research Methods

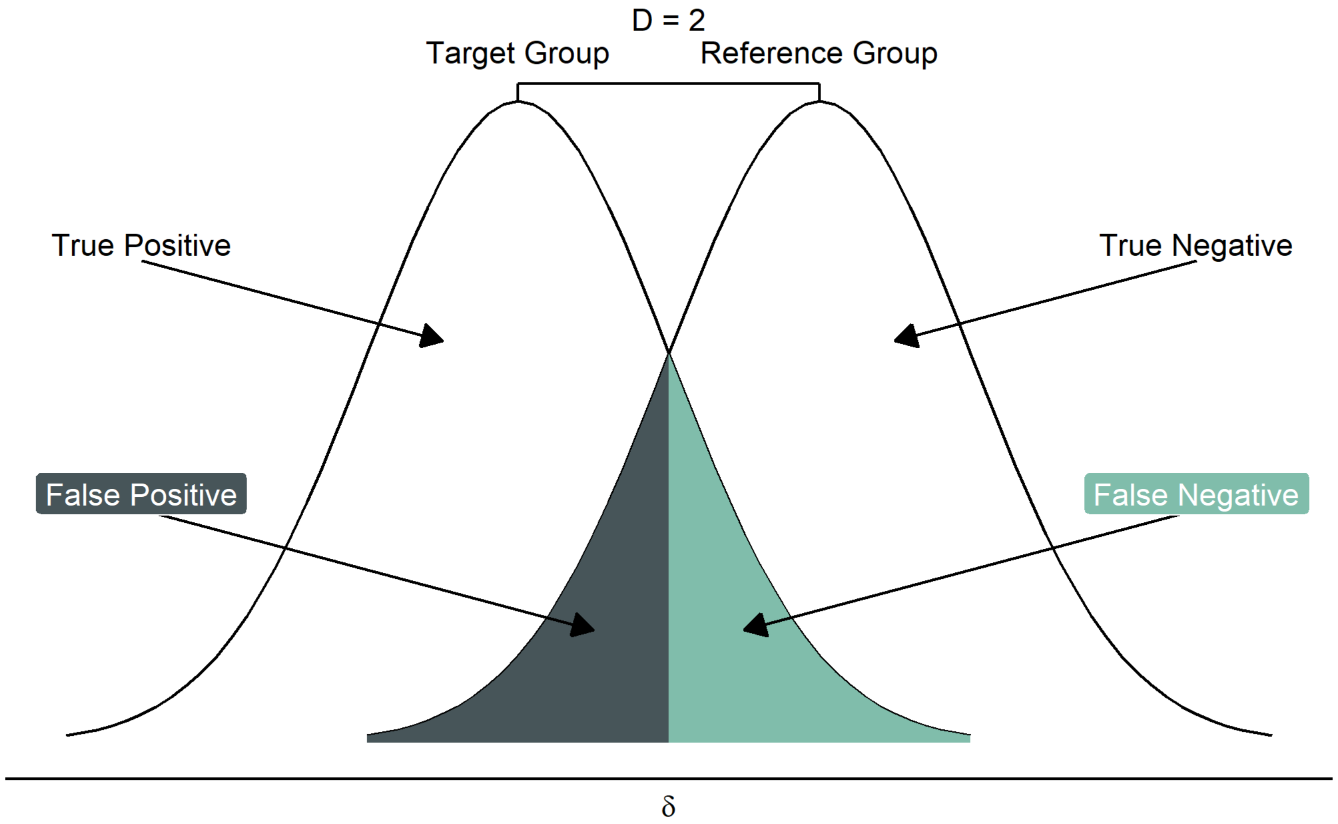
Contact:

| SBGalvin

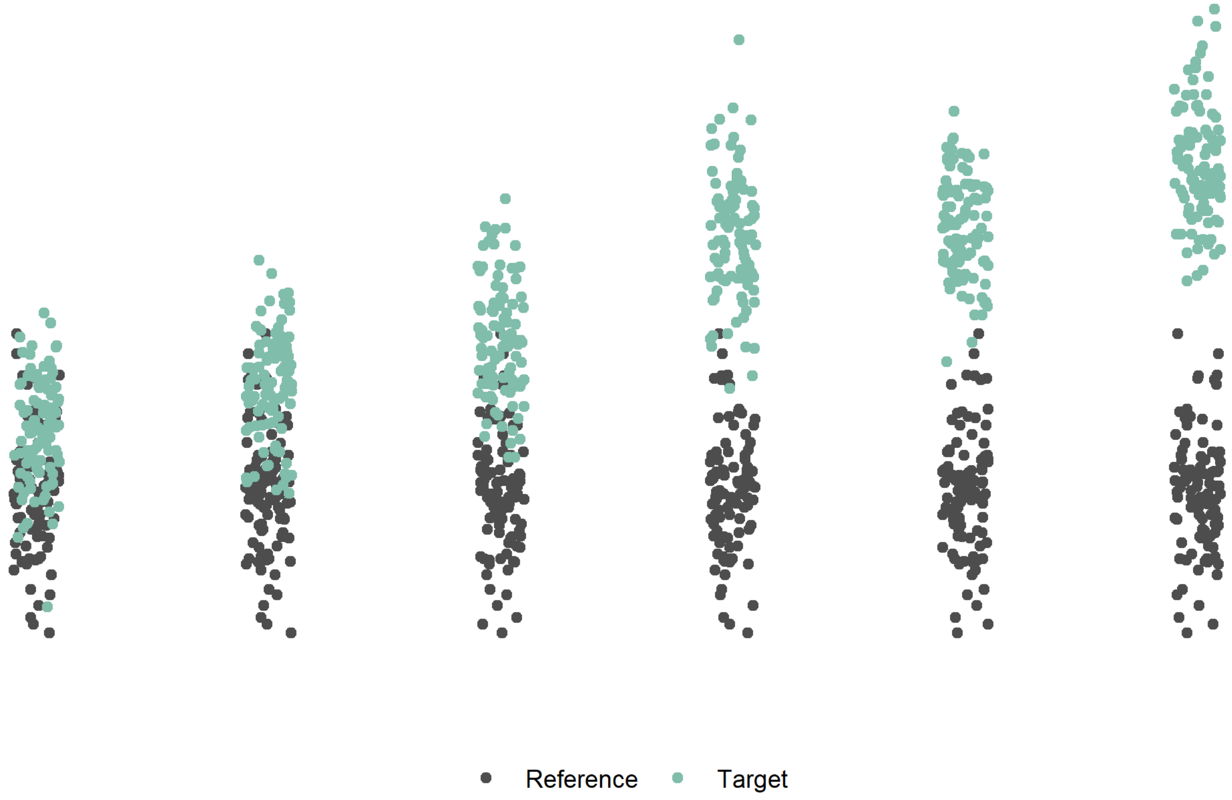
| shane.galvin@ucc.ie

| @S_B_Galvin

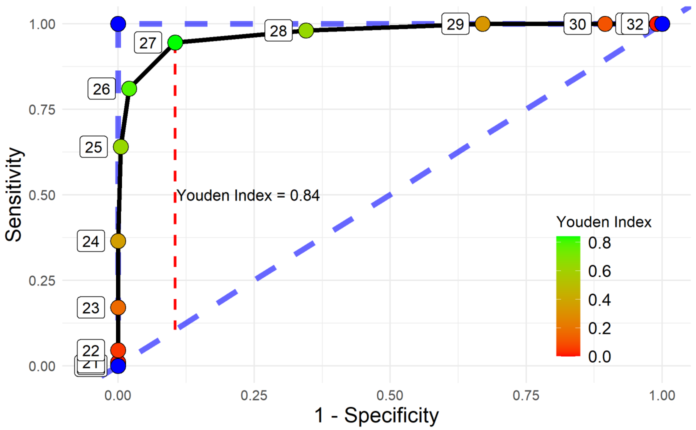
	Desiridatum	Psychometric Topic
Malloy et al., 1997	Less than 15 mins for administration	Test Administration
	Sample all major cognitive domains	Validity
	Demonstrable reliability	Reliability
	Should be able to detect cognitive disorders	Validity
Larner, 2017	Ease of administration	Test Administration
	Ease of interpretation	Test Administration
	Possibility of repeat administration and longitudinal use	Test Construction*

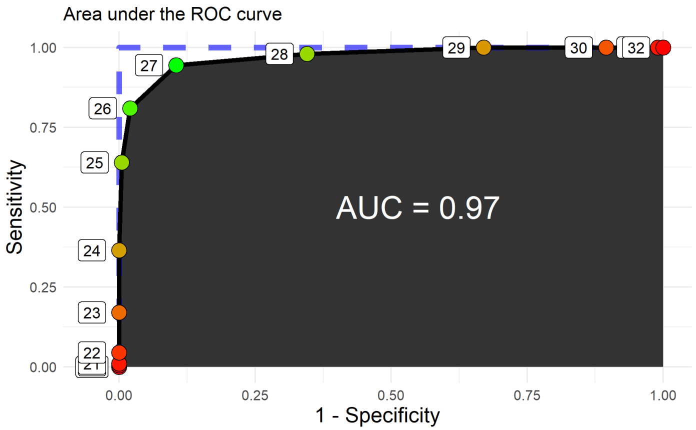

D = 2
Target Group
Reference Group
True Positive
True Negative
False Positive
False Negative
δ

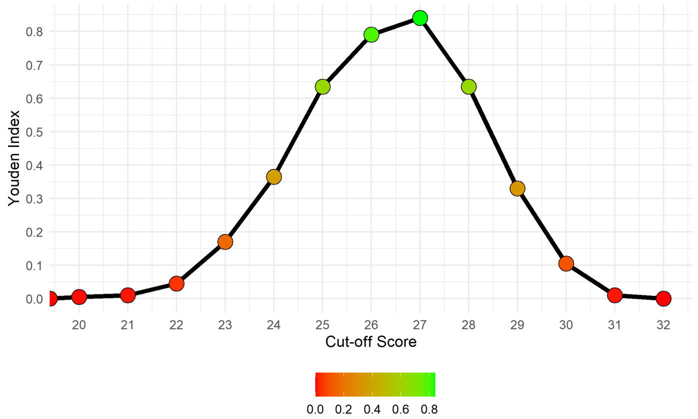
00.4 | Test Construction and Separation

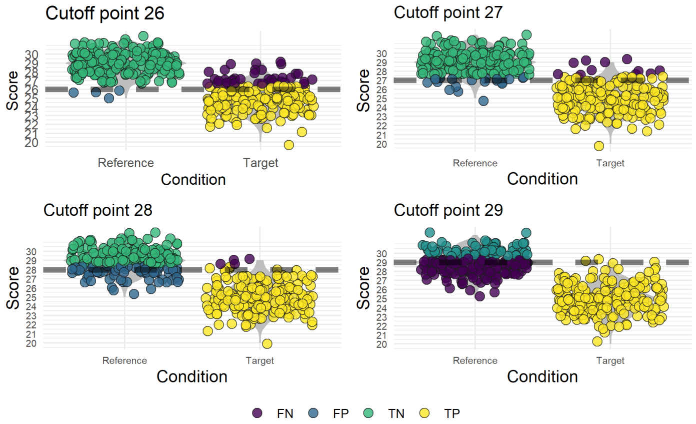


00.5 | Receiver Operating Characteristic curve









00.6 | My Research

Often, CSTs employ classical test theory, which , if we consider Larner and Malloy's CST desiridata, implies the need for:

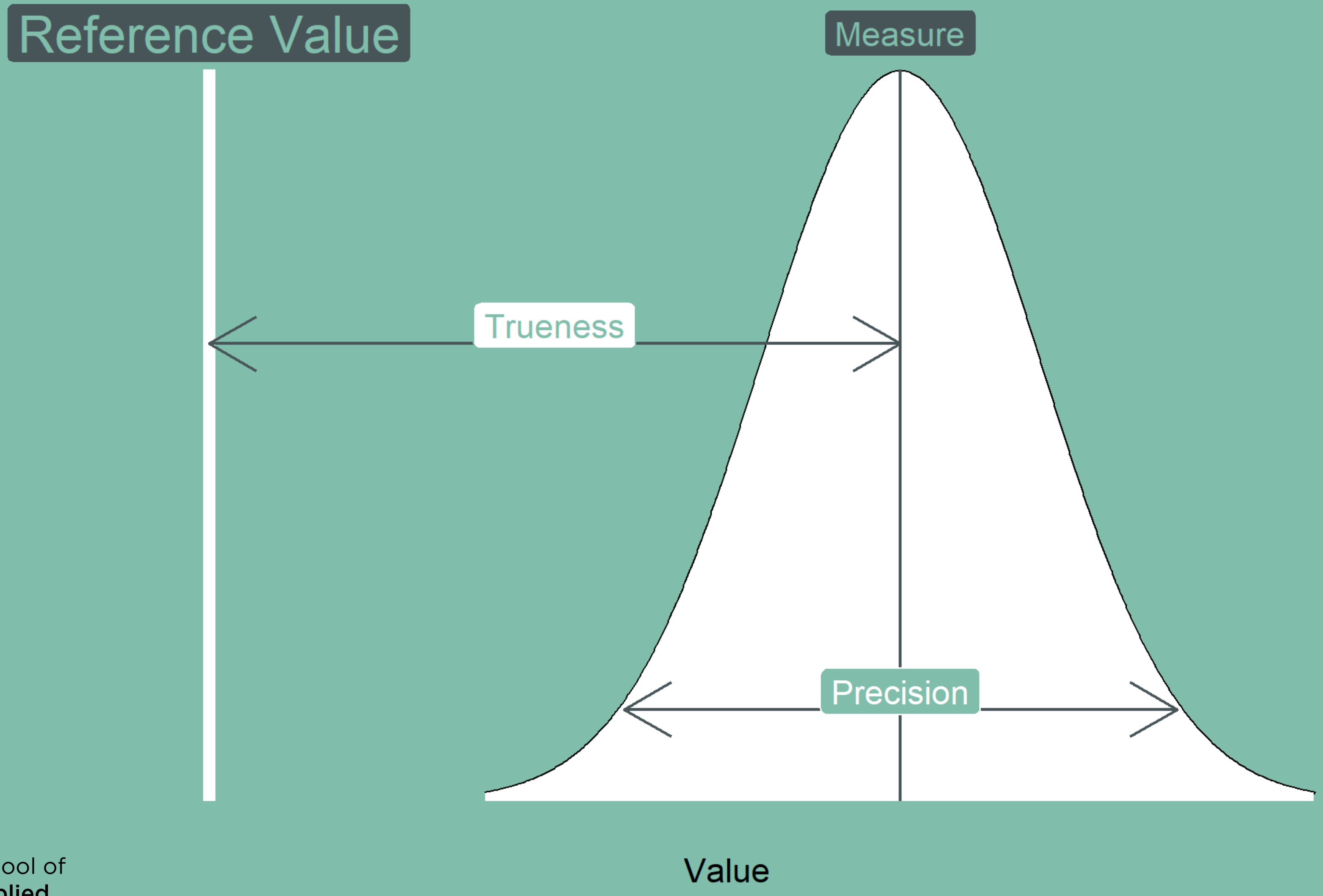
1. Parallel test development (for repeat assessment)
2. High discrimination items (for shortened tests)

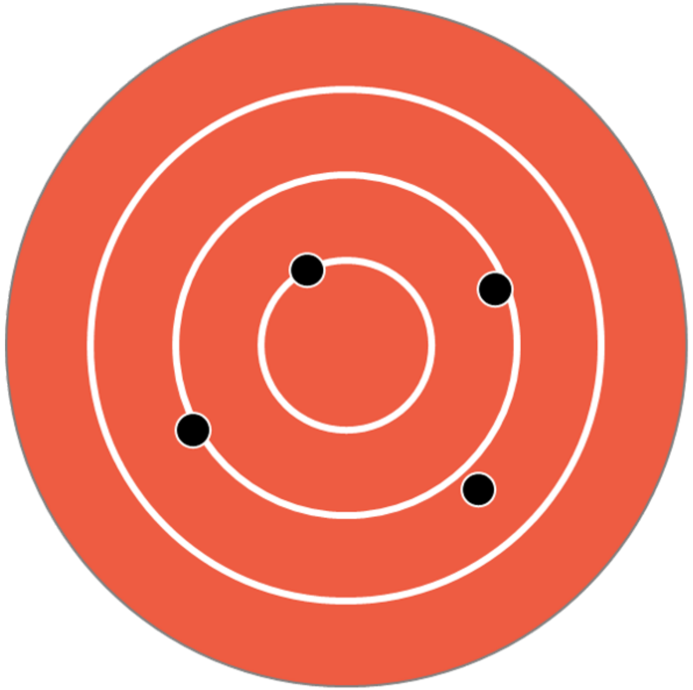
However, few CSTs employ Rasch measurement, or Item Response Theory. There are three advantages either has over CTT:

1. Tests can be calibrated around a given cutoff score, reducing the need for test range.
2. Tests can be administered using Computerised Adaptive Testing.
3. Items can be constructed using Automatic Item Generation.

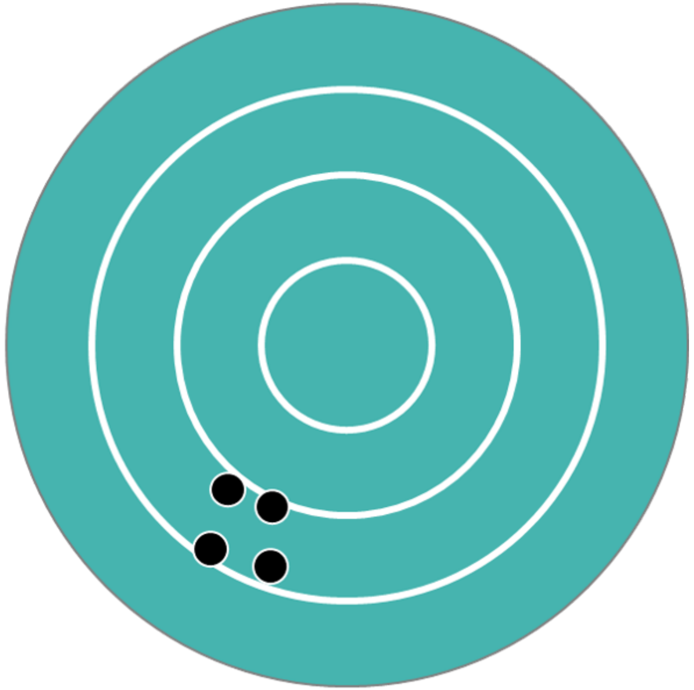
My task is to calibrate a set of items, capable of all of the above, in a healthy sample, and that the calibrated items, and derived items are open-access/free-to-use.

01
Psychological Measurement


Reference Value
Measure
Trueness
Precision
Value



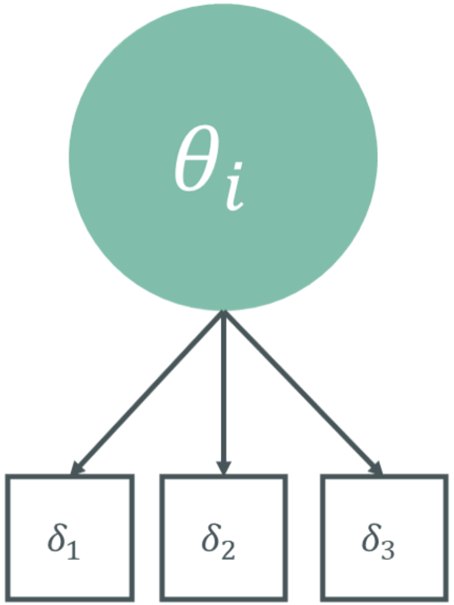
Low Accuracy Due to Poor
Precision



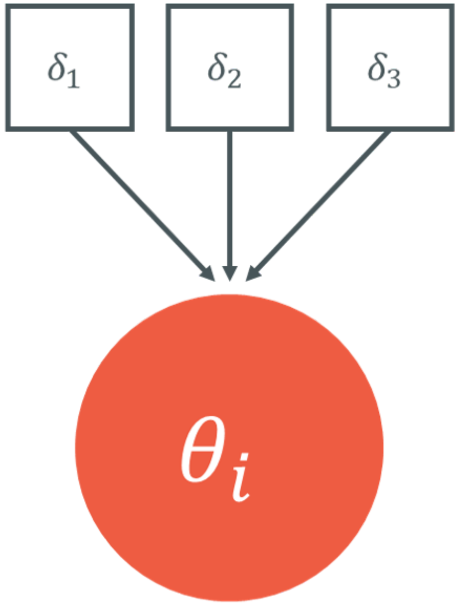
Low Accuracy Due to Poor
Trueness

01.3 | Latent Variable Theory

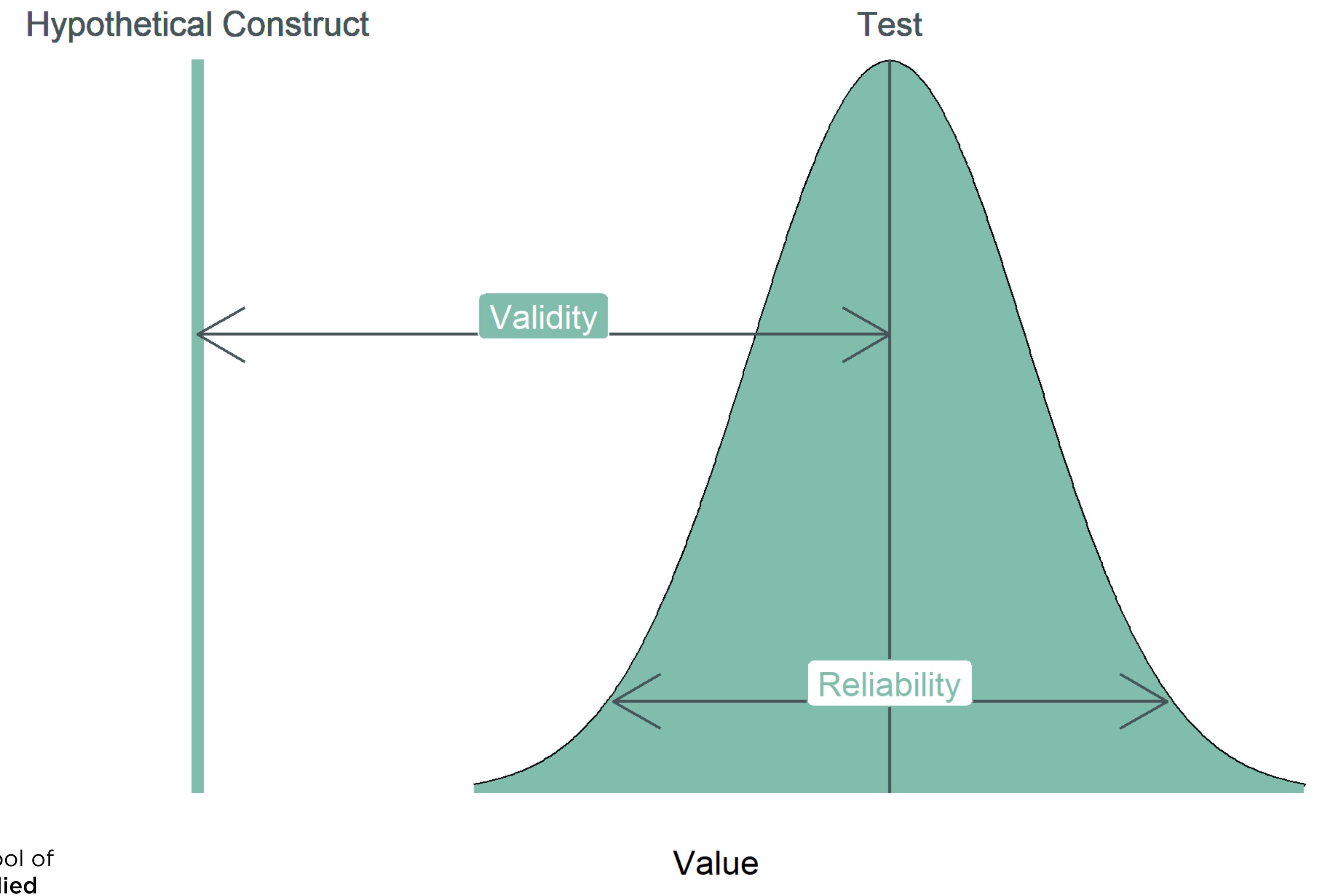
The variables that we are trying to measure are hidden and can only be inferred through statistical connections to observable variables.



Reflective Latent Variable



Formative Latent Variable


Hypothetical Construct
Test
Validity
Reliability
Value

01.5 | Validity and Validation

Validity

In a reflective latent variable model; **variance in the latent variable causes variance in the indicator variable**.

If we take any latent variable θ and any set of indicators δ_i ; δ_i should be **uncorrelated** after controlling for the influence of θ
Validity is a property of a construct

Validation

The process by which we may infer the property of validity. A system of understanding what a test is, or what construct it tends to be belong by relating it to other tests, performing experimental testing, or testing model based predictions.

Validity procedures: e.g. convergent, divergent, predictive Validity

Testing models with pre-specified structure:

Confirmatory Factor Analysis
Rasch Measurement models
Linearised Rasch and IRT models

02
Classical Test Theory

02.1 | True Score Theory

The first true theory of psychological measurement Where X is the observed variable, T its the true score, and E is error

$$X = T + E$$

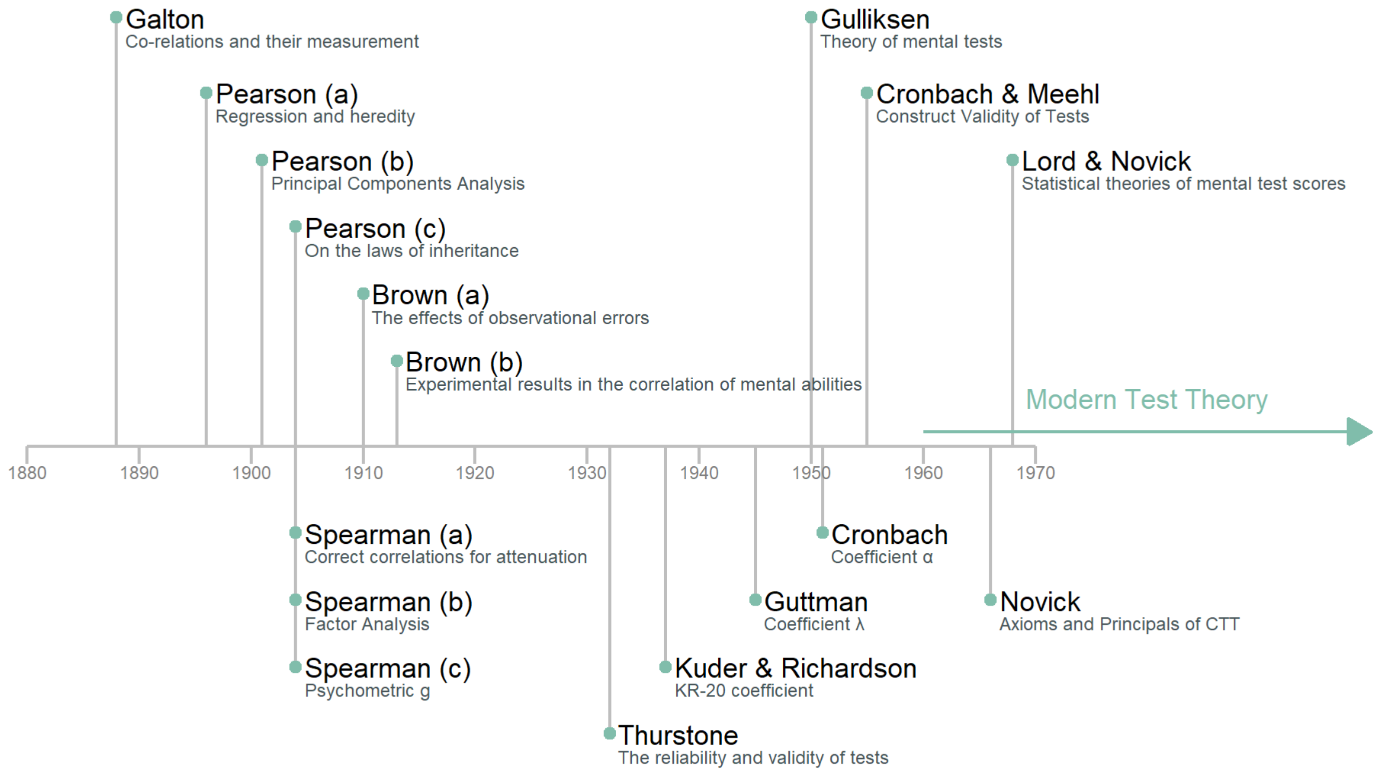
The indicator variable X contains the True score T (θ) Measurement error E is a component of the observed score.

Reliability
Several correlations-type statistics: e.g. KR-20, Gutmann $\lambda_1 - \lambda_6$, Cronbach's α

Validity
Performed via series of contextual tests: e.g. concurrent validity, divergent validity, predictive validity

02.2 | Developments in CTT: Foundations of Psychometrics

Traub (2007)



02.3 | Advantages and Limitations

Classical Test Theory is an enduring approach to modeling psychometric test data, and it's development has provided the language that we need in order to understand, and conduct measurement in a psychological context. However, CTT is also limited by the sample dependent nature of it's formulation and methods (Hambleton and Jones, 1993).

Advantages

(1) A simple linear model links the indicator variable to the latent variable
(2) Easy to compute statistics
(3) Raw score is as simple to understand measure
(4) Relatively relaxed (i.e. weak) model assumptions
(5) Framework for parallel test development

Limitations

(1) Test scores are sample dependent
(2) Item statistics are sample dependent
(3) Whole-test focus means longer test forms
(7) Theory driven test development and confirmatory testing is difficult due to these variances:

(i) Standard error is constant across a test
(ii) Reliability is constant across a test
(iii) Item indices may be volatile

03
Rasch Measurement

A measurement method that defines the relationship between a latent variable and observed score as function of an item location and person ability, making the statement that the greater the persons ability, the higher their probability of a correct response. Rasch models present the structure data should satisfy in order to obtain measurements

Positives

(1) The relationship between a latent variable and observed score as function of an item location and person ability
(2) Item and Ability locations can be estimated independently
(3) Unidimensional model (items should not be related after conditioning on the latent variable)
(4) Standard error and reliability are variable across the `$\theta$` range
(5) Resistant to the shape of the raw score distribution (uniform is best using CMLe)

Negatives

(1) Rasch models can be difficult to fit to real data
(2) Intensive procedure for validating and exploration
(3) `$\chi^2$` fit statistics can be unstable

03.2 | The Dichotomous Rasch Model

The Dichotomous Rasch model takes a set of scored data and produces a conditional probability of correct response for a person to an item.

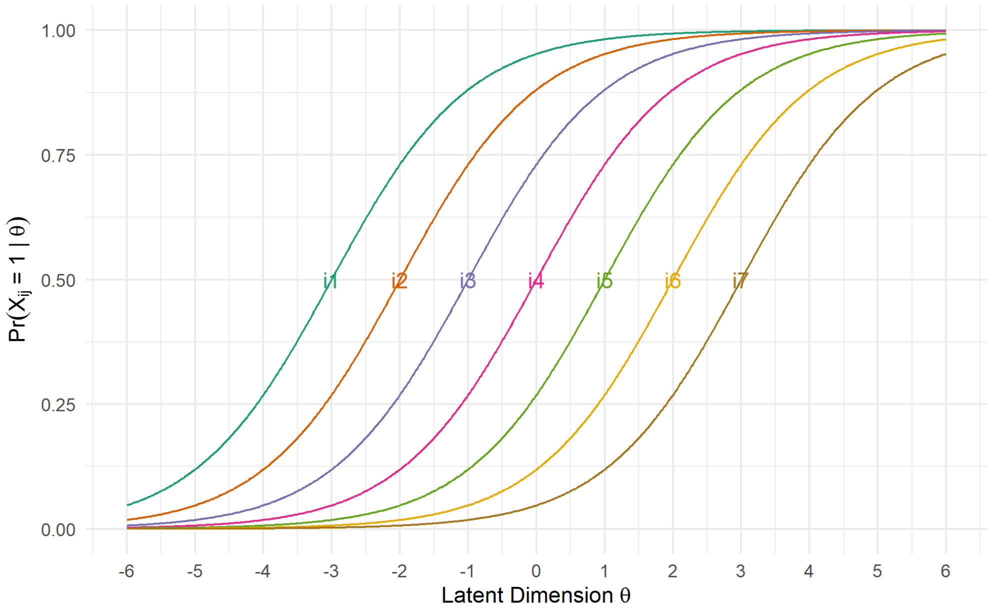
Where:

- $P(X_{ij} = 1|\theta_i\beta_j)$ is the conditional probability of a correct response by a person i to an item j
- θ is the person ability position on the latent scale
- β is the item location on the latent scale

$$P(X_{ij} = 1|\theta_i\beta_j) = \frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}}$$

Rasch Item Parameters

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
ß	-3	-2	-1	0	1	2	3



School of Applied Psychology

03.3 | Partial Credit Rasch Model

The Partial Credit Model (Masters, 1982) is a generalisation of the dichotomous Rasch model for items with 2 or more ordered outcome categories. When there are two categories, the PCM is equivalent to the dichotomous model. A unique advantage of the PCM is that items can have their own scoring sytem (i.e. number of categories).

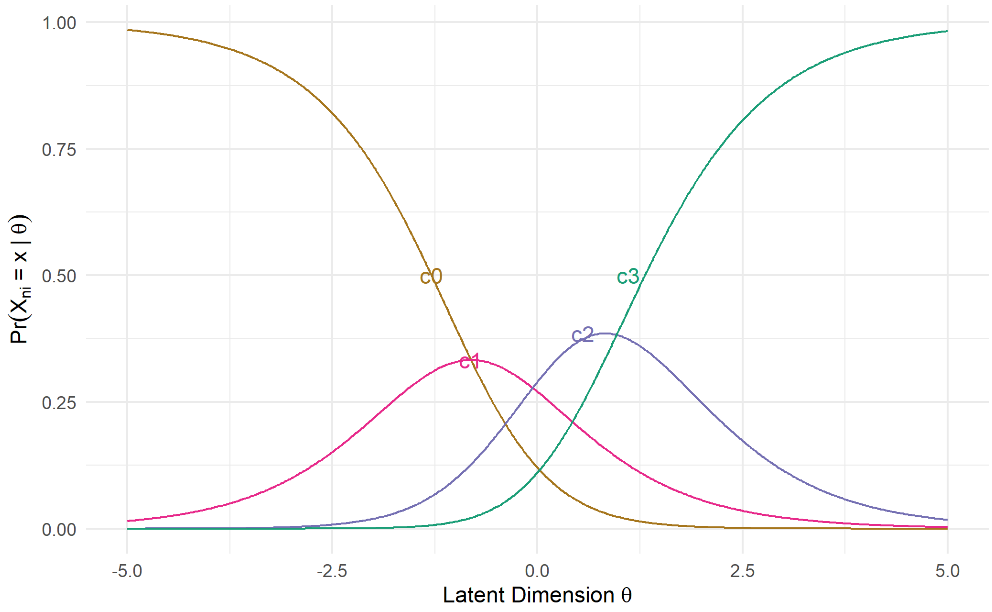
Where:

- $P(X_{ni} = x)$ is the probability of selecting an outcome x for person n to item i
- τ_{ki} is the k^{th} threshold location of item i on the latent scale (denoted c in the table and plot)
- θ_n is the person ability position on the latent scale
- m_i is the maximum score for item i

$$P(X_{ni} = x) = \frac{exp\sum_{k=0}^{x}(\theta_n - \tau_{ki})}{\sum_{j=0}^{m_i} exp\sum_{k=0}^{j}(\theta_n - \tau_{ki})}$$

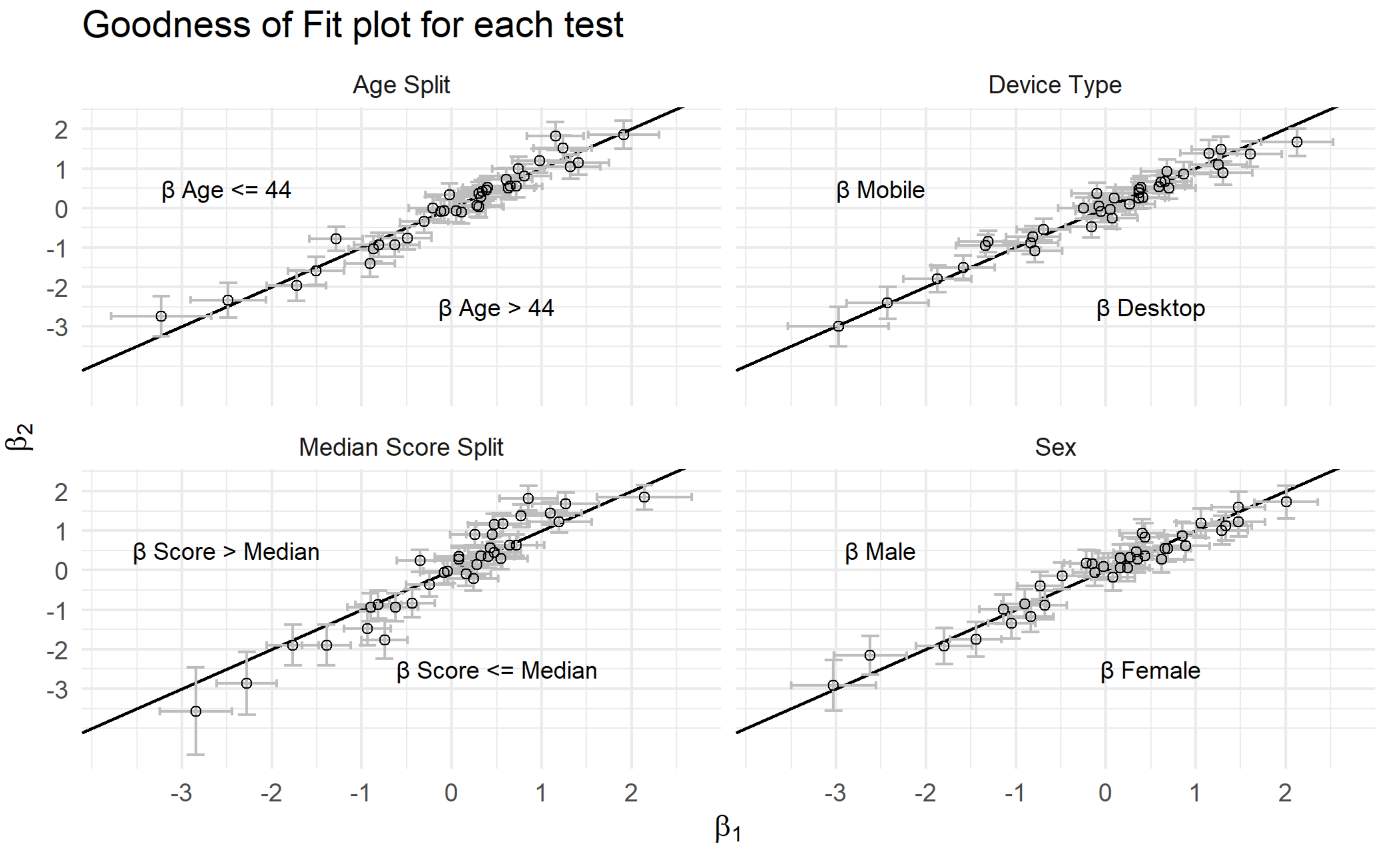
Threshold Parameters

Item	β	c1	c2	c3
i3	0.024	-0.811	0.066	0.816



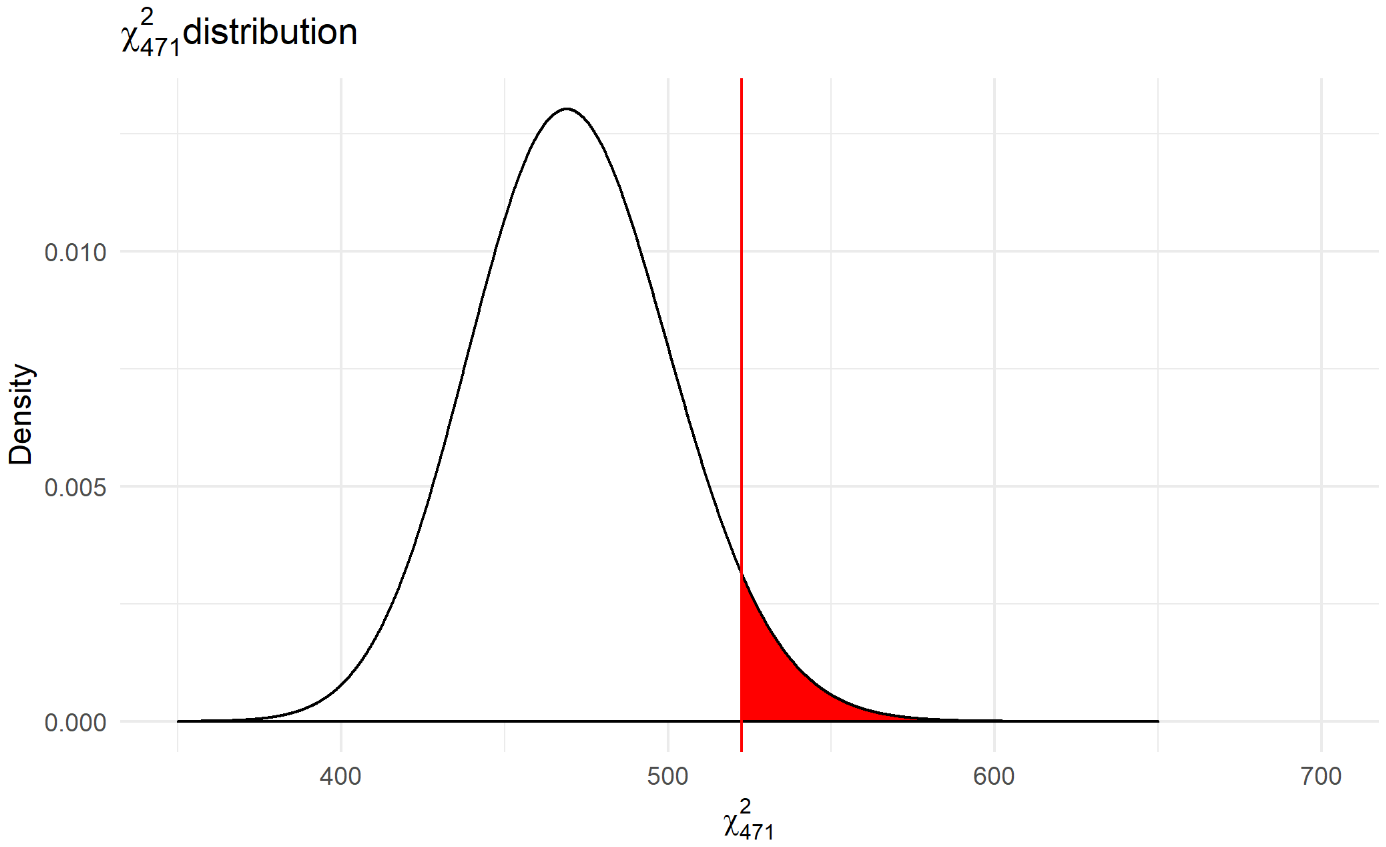
03.4 | Model Fit Statistics

Andersen's Likelihood Ratio test is commonly used as a global fit test. and is based on the ratio of the likelihood of the response pattern based on a single overall ability estimate, and the likelihood of the response pattern based on the estimated abilities for each subset. When the subsets have the same ability parameters the likelihood ratio is χ^2 distributed with degrees of freedom = n items - 1.



03.5 | Item Fit Statistics

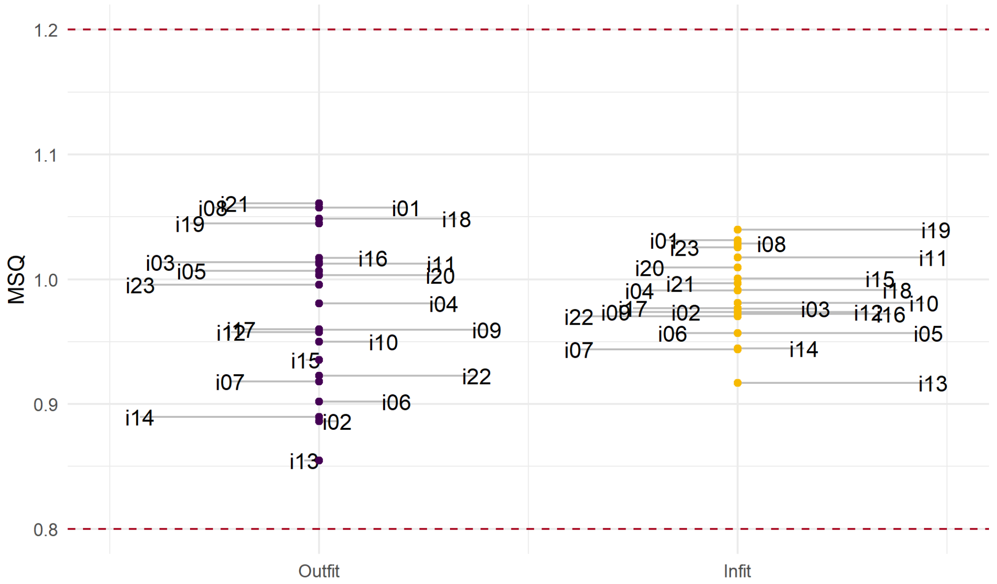
Item fit statistics are based on Rasch residuals and further computations which are designed to assess the fit of an item to the model. The are assumed to be χ^2 distributed with N = number of participants -1



03.6 | Infit and Outfit Mean Square

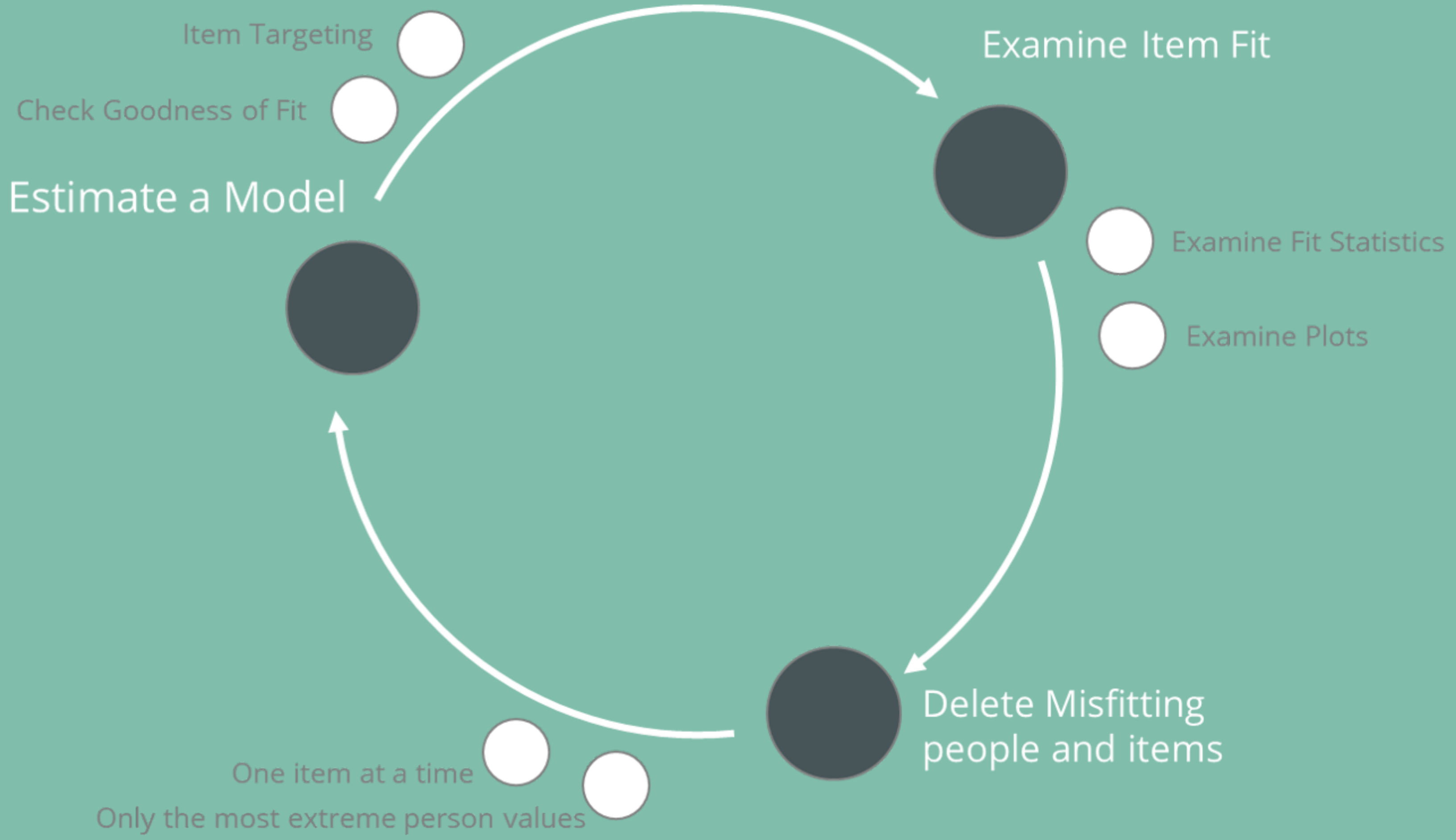
Infit and Outfit illustrate the amount of randomness or distortion in measurement.

- Infit:
  - A weighted statistic, sensitive to targeting of a person (overfit for Guttman patterns)
- Outfit:
  - An unweighted statistic; Sensitive to items with difficulty distant from ability.
  - Overfit for imputed missing data, underfit for guessing patterns.
- Caveat:
  - because outfit values are calculated $\frac{\chi^2}{df}$
  - as sample size increases the power to detect small effects also increase
  - making small departures from 1.0 statistically significant
  - increasing type 1 error rate.



04
Item Investigation

04.1 | Exploratory Rasch Analysis Cycle



04.2 | How do you investigate the item for partial credit scoring?

(A) Table Method (Andrich and Styles, 2009)

(1) Divide the person parameter `$\theta$` into 3 class intervals
(2) Produce a proportion value for each class interval for each response option
(3) Establish a chance level of response selection
(4) Compare the proportions in each category to the chance level
(5) Options with proportions higher than chance level should be considered for partial credit
(6) If only the correct option has a proportion and the remaining outcomes are below the chance level then a dichotomous model may be best

(B) Graphical Method (Asril and Marais, 2011)

(1) Divide the person parameter `$\theta$` into n class intervals
(2) Produce a proportion value for each class interval for each response option
(3) Plot the empirical response curves (ERC) against the Rasch dichotomous IRF
(4) If all ERC are monotonic non-increasing then a dichotomous model may be best
(5) If any ERC has a peak or is non-decreasing; consider awarding partial credit

(C) Examine Item and compare the overlap between distractor options and the correct option!

04.3 | How to do this in R with psketti

psketti

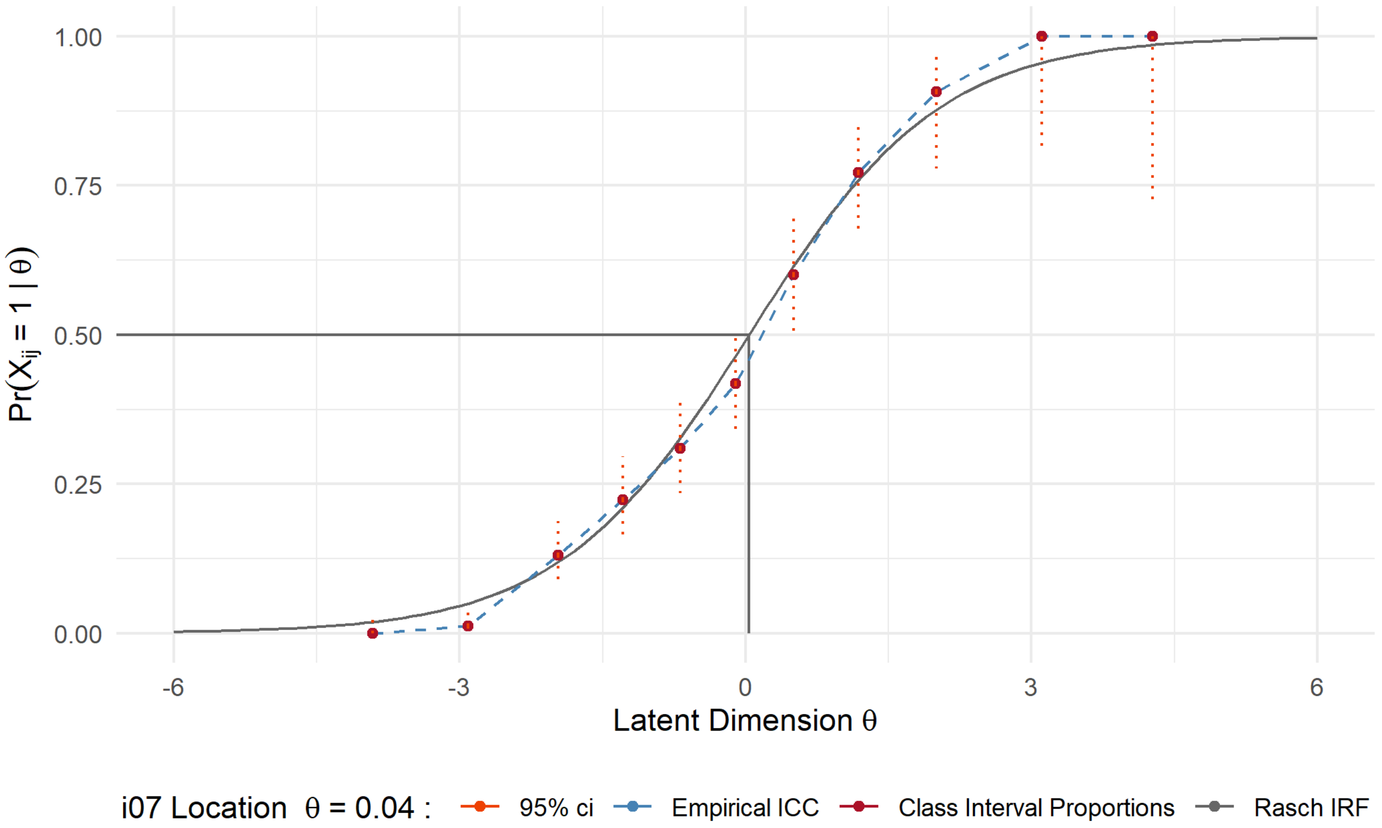
A package to generate investigatory Rasch model plots. Package developed with reference to Andrich and Styles (2009), and Asril and Marais (2011).

```
# install.packages("devtools") # install devtools ## run once
# devtools::install_github("SBGalvin/psketti")  # install psketti ## run once
options(scipen = 999, stringsAsFactors = FALSE) # options
# Load Packages -----
library(eRm)       # for estimating Rasch Models
library(psketti)   # graphical investigation
data("FakeData")   # load fake data from psketti
# set up data for eRm
Fake_Data_scores <- reshape(FakeData[, c("ID", "Item", "X")],
                            timevar = "Item",
                            idvar = "ID",
                            direction = "wide")

names(Fake_Data_scores) <- c("ID", paste0("i", sprintf(fmt  = "%02d", 1:23)))
row.names(Fake_Data_scores) <- Fake_Data_scores$ID # apply ID code to row names
Fake_Data_scores$ID <- NULL # drop ID code as a column

fake_rm <- RM(Fake_Data_scores) # estimate dichotomous Rasch model
```

```
# extract data from eRm estimated Rasch model ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
psk_data  <- pskettify(fake_rm)   # pskettify

# plot data ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
psk_IRF   <- psketti(psk_data)    # IRF plots
psk_IRF$Plot.List[['i02']][[1]]   # prints to plot viewer

# Styles-Andrich table ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

r_o <- factor(sort(unique(FakeData$K)),                # Response options as factor
              levels = sort(unique(FakeData$K)),       # factor levels
              ordered = TRUE)                          # ordered factor

tlt_data <- tabliatelle(x = FakeData,                  # data object
                        eRm.obj = fake_rm,             # estimated Rasch model
                        ID = "ID",                     # ID column
                        Item = "Item",                 # Item column
                        K = "K",                       # Category column
                        response_options = r_o)        # response options


spag_plot <- psketti_distractor(ID = "ID",                  # set ID column
                                Item = "Item",              # set Item column
                                K= "K",                     # Set resp categories
                                x = FakeData,               # select data
                                eRm.obj = fake_rm,          # select eRm object
                                response_options = r_o,     # set response options
                                p.style = "present")        # set plotting style

spag_plot$Plot.List[['i01']][[1]] # plot s
```

04.4 | Rasch Empirical and Theoretical IRF

The empirical curve (blue) is plotted against the Rasch IRF, with empirical proportions (dark red) for ability class intervals and their 95% confidence intervals (dotted red lines).



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh
School of Applied Psychology

04.5 | Class-Interval Table

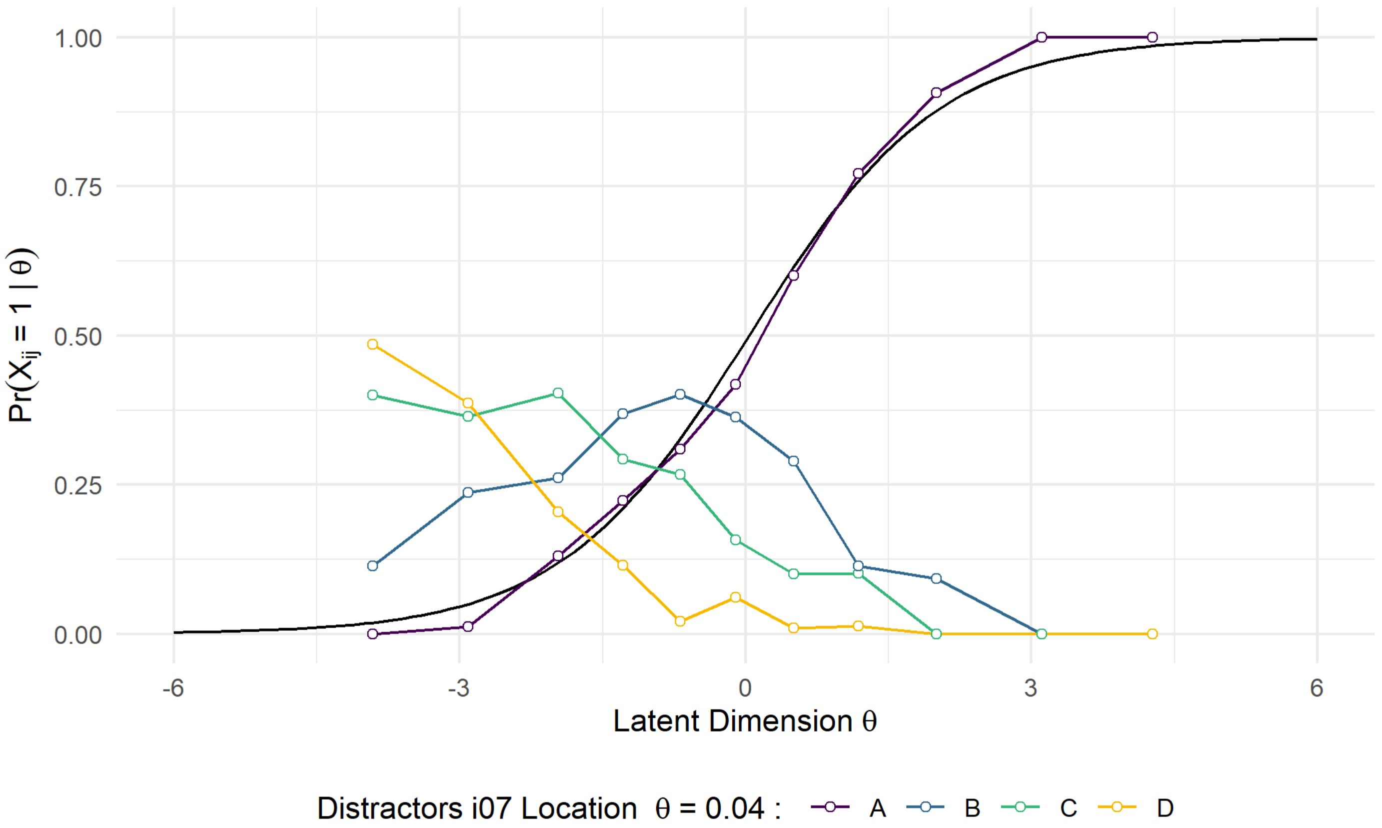
IF the dichotomous model is the true (better) model; then the class proportions for each distractor options should not exceed a value of:

$$p = \frac{1}{N_{distractor\ options}}$$

Andrich-Styles Table: i07

Item	β	Class Interval	θ	00	01	02	03
i07	0.04	1	-2.46	0.29	0.37	0.25	0.09
		2	0.07	0.03	0.17	0.32	0.48
		3	2.63	0.00	0.00	0.06	0.94

04.6 | Comparing Dichotomous Scoring with Raw MCQ Categories



04.7 | Examining Item Content

Initially, we deemed there to be an objectively correct answer to a Multiple Choice question; so we estimated a dichotomous Rasch Model. If this does not fit as well as we hoped, and our Class interval table and distractor plot highlights some irregularities, then we should examine the item content.

If you have 4 pencils and I have 7 apples, how many pancakes will fit on the roof?

1. PANCAKES!!!!!????!!!
2. 2193.266
3. 28
4. Purple, because aliens don't wear hats.

Despite there being a clearly correct answer; some distractor options may overlap with the correct option. In this case we can assign partial credit:

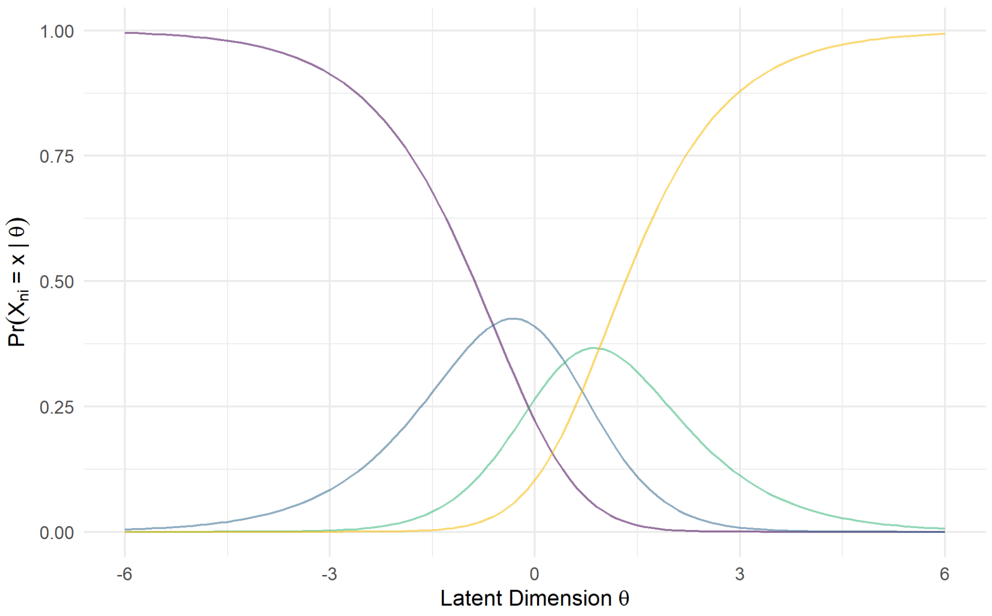
1. PANCAKES!!!!!????!!! <- Score category 2
2. 2193.266 <- Score category 0
3. 28 <- Score category 0
4. Purple, because aliens don't wear hats. <- Score category 1

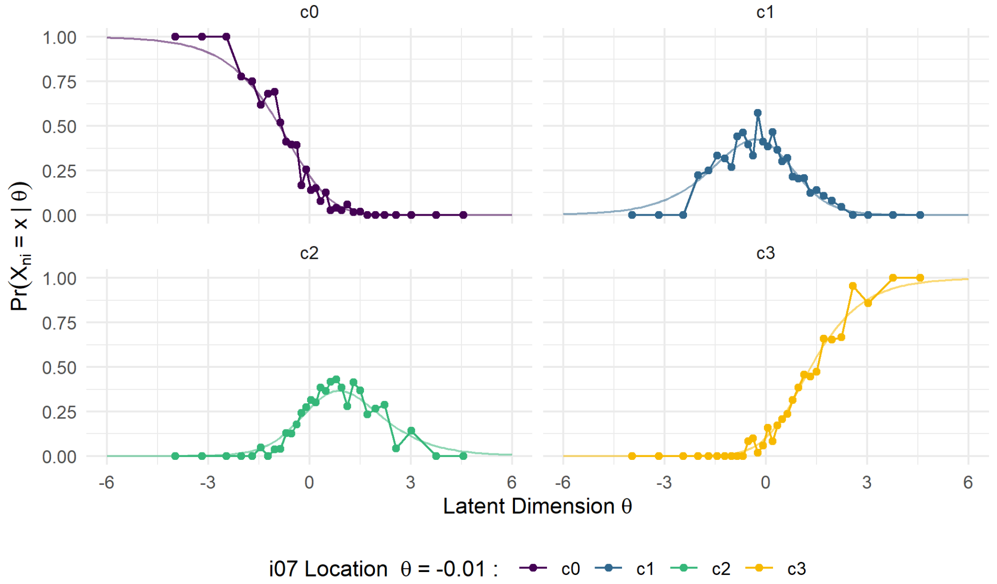
UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh | School of Applied Psychology

04.8 | Result of Fitting a PCM

Partial Credit Model Item Fit Statistics

Item	K	tau	Se	χ^2	df	Outfit	Infit
i07	c1	-0.61	0.09	1090.34	1184	0.92	0.93
i07	c2	-0.18	0.10				
i07	c3	0.76	0.12				





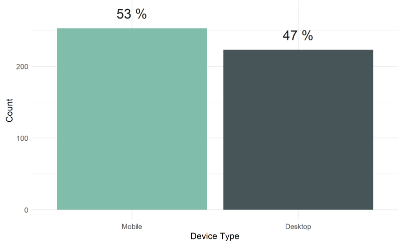
05

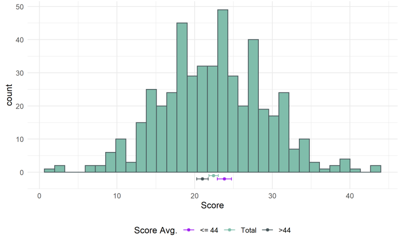
The MaRs IB

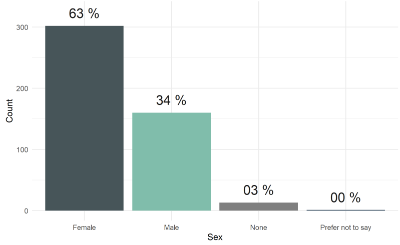
05.1 | Project Manchego: Description

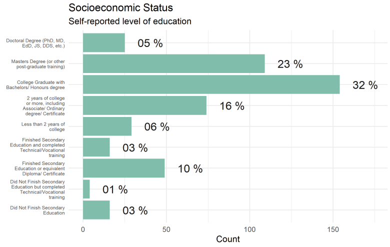
Sample Age, broken down by Biological Sex

		N	Mean	Median	St.Dev	Min	Max
Total		476	43.35	44	43.35	18	78
Sex	Female	302	43.68	45	43.68	18	78
	Male	160	41.66	38	41.66	18	72
	None	13	58.54	60	58.54	39	73
	Prefer not to say	1	18.00	18	18.00	18	18

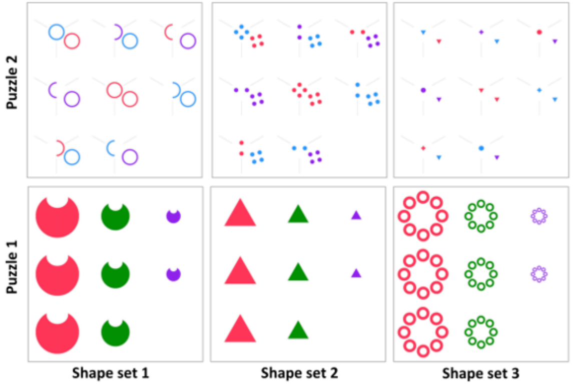








Example MaRs Items

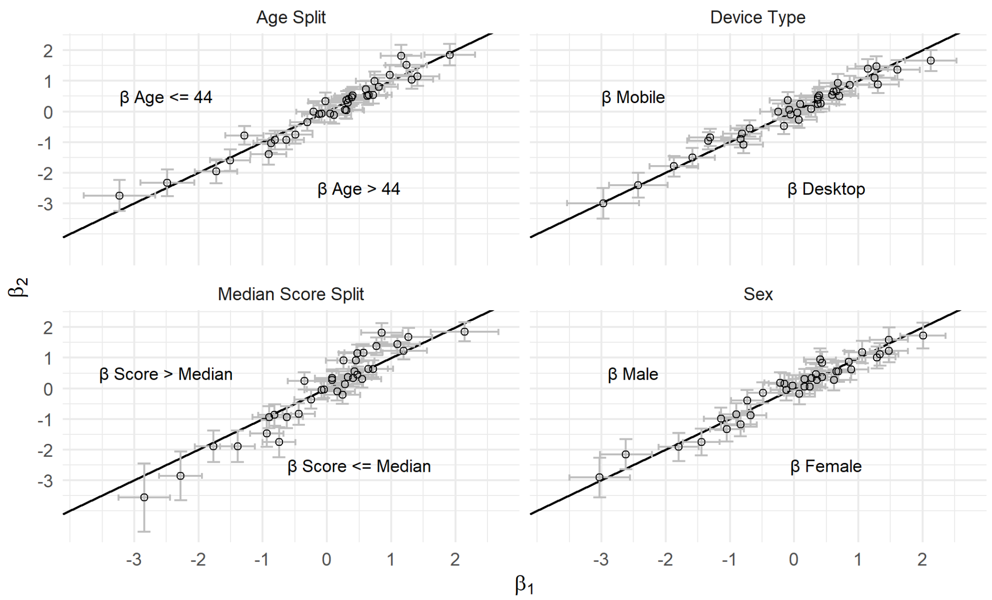


Design

- 45 items administered in pseudo-randomised order
- 30 second time limit per item
- Incomplete cases not recorded
- Raw Responses, RT, Scored Response (correct/incorrect)

05.2 | Project Manchego: Rasch Analysis

Test Criteria	LR Value	χ² df	p
Median Score Split	112.91	35	0.0000000
Age Split	42.28	35	0.1855833
Device Type	37.40	35	0.3594916
Sex	41.03	35	0.2230208

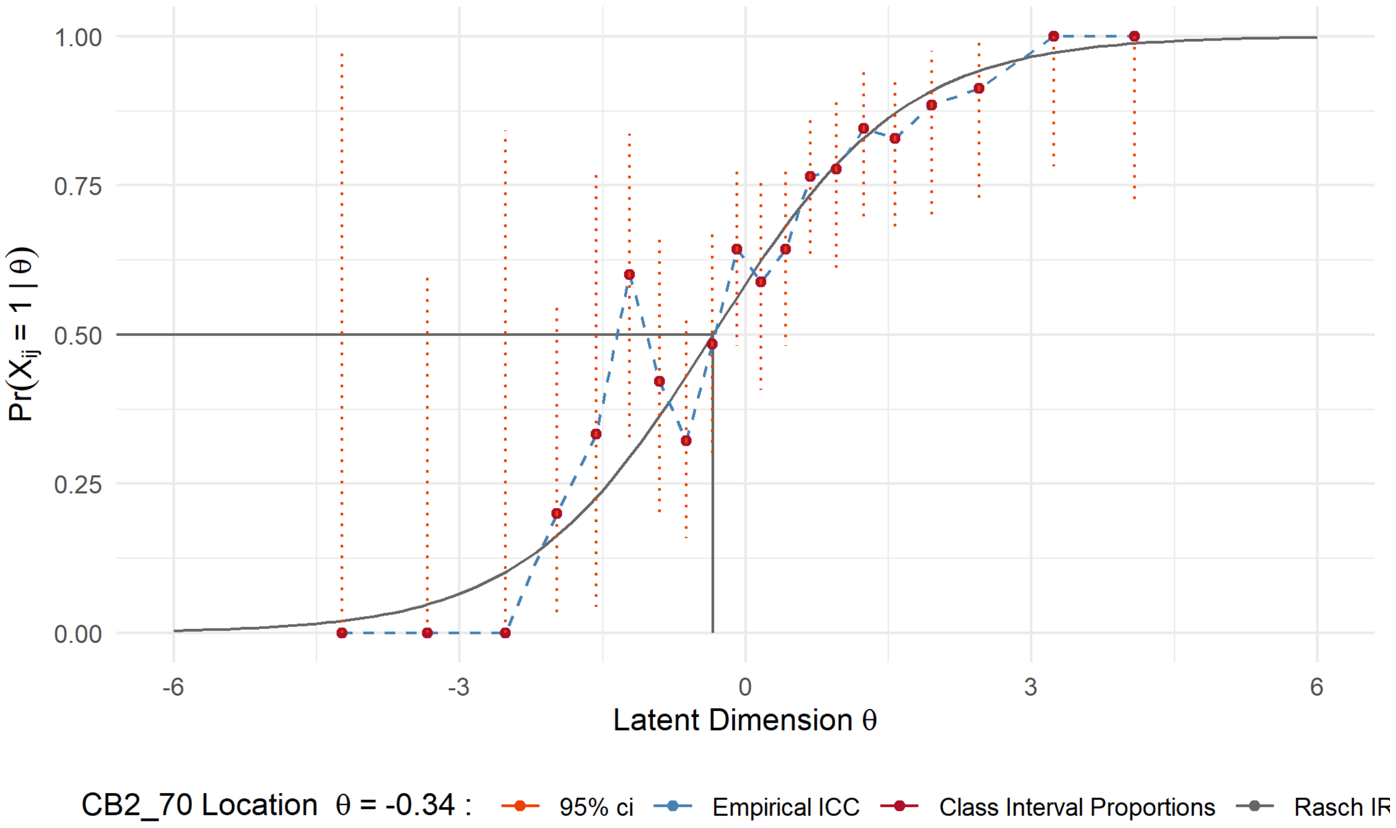


05.3 | Project Manchego: Item Fit Statistics

Item	β	Se	Discrimination	χ^2	df	p	Outfit	t	Infit	t
CB2_06	0.00	0.10	0.37	449.96	455	0.56	0.99	-0.23	0.97	-0.74
CB2_07	-2.99	0.19	0.24	342.03	455	1.00	0.75	-1.01	0.97	-0.16
CB2_08	-2.41	0.15	0.31	350.60	455	1.00	0.77	-1.30	0.90	-0.90
CB2_09	-1.83	0.13	0.30	418.43	455	0.89	0.92	-0.60	0.97	-0.38
CB2_10	-0.12	0.10	0.20	536.58	455	0.00	1.18	3.13	1.11	3.24
CB2_11	1.37	0.12	0.15	538.89	455	0.00	1.18	1.62	1.08	1.27
CB2_12	1.88	0.13	0.34	372.57	455	1.00	0.82	-1.29	0.94	-0.73
CB2_13	1.08	0.11	0.20	524.57	455	0.01	1.15	1.61	1.08	1.51
CB2_15	0.81	0.11	0.23	502.93	455	0.06	1.10	1.33	1.07	1.44
CB2_16	0.14	0.10	0.29	476.22	455	0.24	1.04	0.81	1.04	1.17
CB2_17	0.46	0.10	0.36	444.44	455	0.63	0.97	-0.39	0.98	-0.61
CB2_18	0.57	0.10	0.26	476.53	455	0.23	1.05	0.69	1.06	1.41
CB2_19	-0.11	0.10	0.32	482.15	455	0.18	1.06	1.07	1.01	0.35
CB2_20	-1.12	0.11	0.37	403.44	455	0.96	0.88	-1.38	0.95	-1.03
CB2_22	-0.77	0.10	0.42	420.74	455	0.87	0.92	-1.13	0.91	-2.21
CB2_23	-0.33	0.10	0.38	448.05	455	0.58	0.98	-0.30	0.97	-0.95
CB2_25	-0.86	0.10	0.33	433.05	455	0.76	0.95	-0.68	1.00	0.01
CB2_26	1.26	0.11	0.28	460.73	455	0.42	1.01	0.14	1.00	0.00

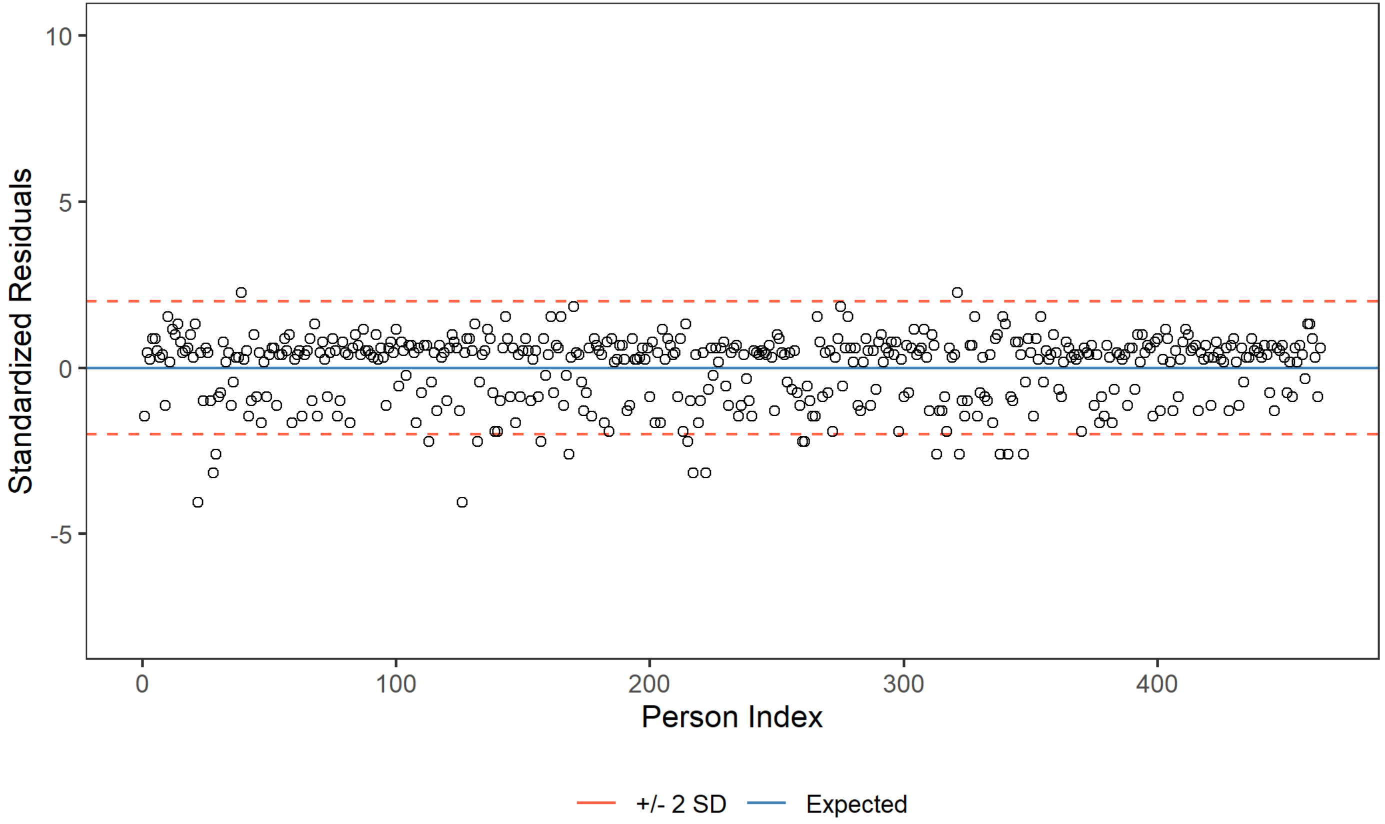
Item	β	Se	Discrimination	χsup2	df	p	Outfit	t	Infit	t
CB2_27	0.42	0.10	0.31	480.00	455	0.20	1.05	0.87	1.02	0.42
CB2_28	-0.08	0.10	0.35	443.06	455	0.65	0.97	-0.52	0.99	-0.34
CB2_29	0.30	0.10	0.29	503.17	455	0.06	1.10	1.74	1.03	0.86
CB2_30	0.38	0.10	0.40	421.67	455	0.87	0.92	-1.27	0.95	-1.35
CB2_35	0.34	0.10	0.39	428.78	455	0.81	0.94	-1.02	0.96	-1.20
CB2_39	0.17	0.10	0.29	464.01	455	0.37	1.02	0.33	1.04	1.05
CB2_40	0.66	0.10	0.27	467.79	455	0.33	1.03	0.39	1.05	1.11
CB2_45	0.17	0.10	0.43	422.89	455	0.86	0.93	-1.32	0.93	-2.15
CB2_49	-1.54	0.12	0.41	378.03	455	1.00	0.83	-1.60	0.87	-2.06
CB2_50	-0.62	0.10	0.45	399.79	455	0.97	0.88	-2.01	0.91	-2.52
CB2_55	0.60	0.10	0.41	403.81	455	0.96	0.89	-1.75	0.94	-1.46
CB2_60	0.87	0.11	0.17	551.05	455	0.00	1.21	2.50	1.10	2.08
CB2_65	-1.05	0.11	0.51	356.46	455	1.00	0.78	-2.87	0.85	-3.26
CB2_69	-0.01	0.10	0.47	397.02	455	0.98	0.87	-2.50	0.90	-2.93
CB2_70	-0.94	0.11	0.33	439.42	455	0.69	0.96	-0.46	0.98	-0.35
CB2_75	1.17	0.11	0.31	438.46	455	0.70	0.96	-0.38	0.99	-0.13
CB2_79	1.48	0.12	0.18	529.60	455	0.01	1.16	1.37	1.08	1.13
CB2_80	0.63	0.10	0.35	452.53	455	0.52	0.99	-0.09	0.99	-0.13

05.4 | Item 70 IRF



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh
School of Applied Psychology

05.5 | Item 70 Residuals



05.6 | Item 70 Class Interval Table

IF the dichotomous model is the true (better) model; then the class proportions for each distractor options should not exceed a value of:

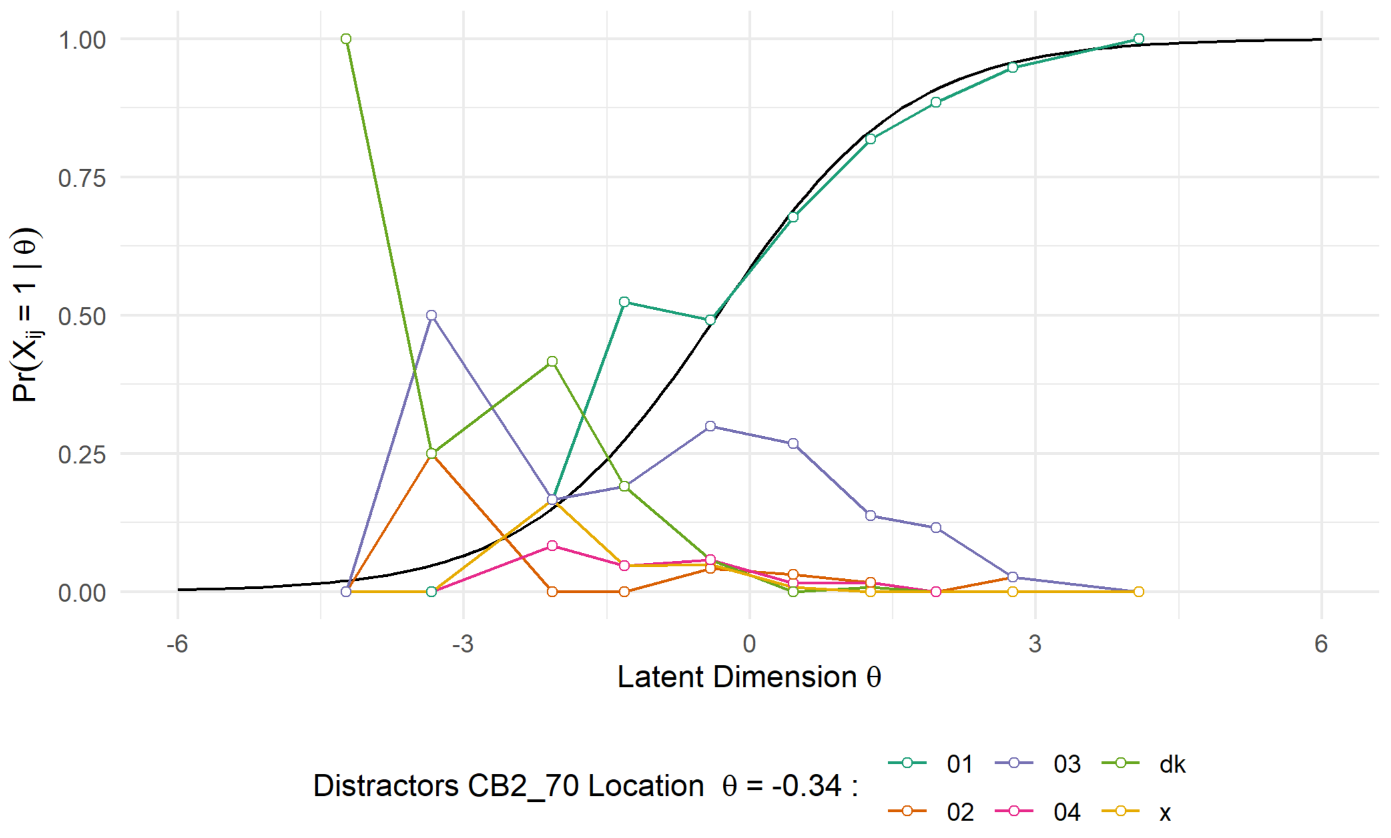
$$p = \frac{1}{N_{distractor\ options}}$$

IF the class proportion exceeds p , then the item should be noted for further investigation.

Andrich-Styles Table: CB2 70

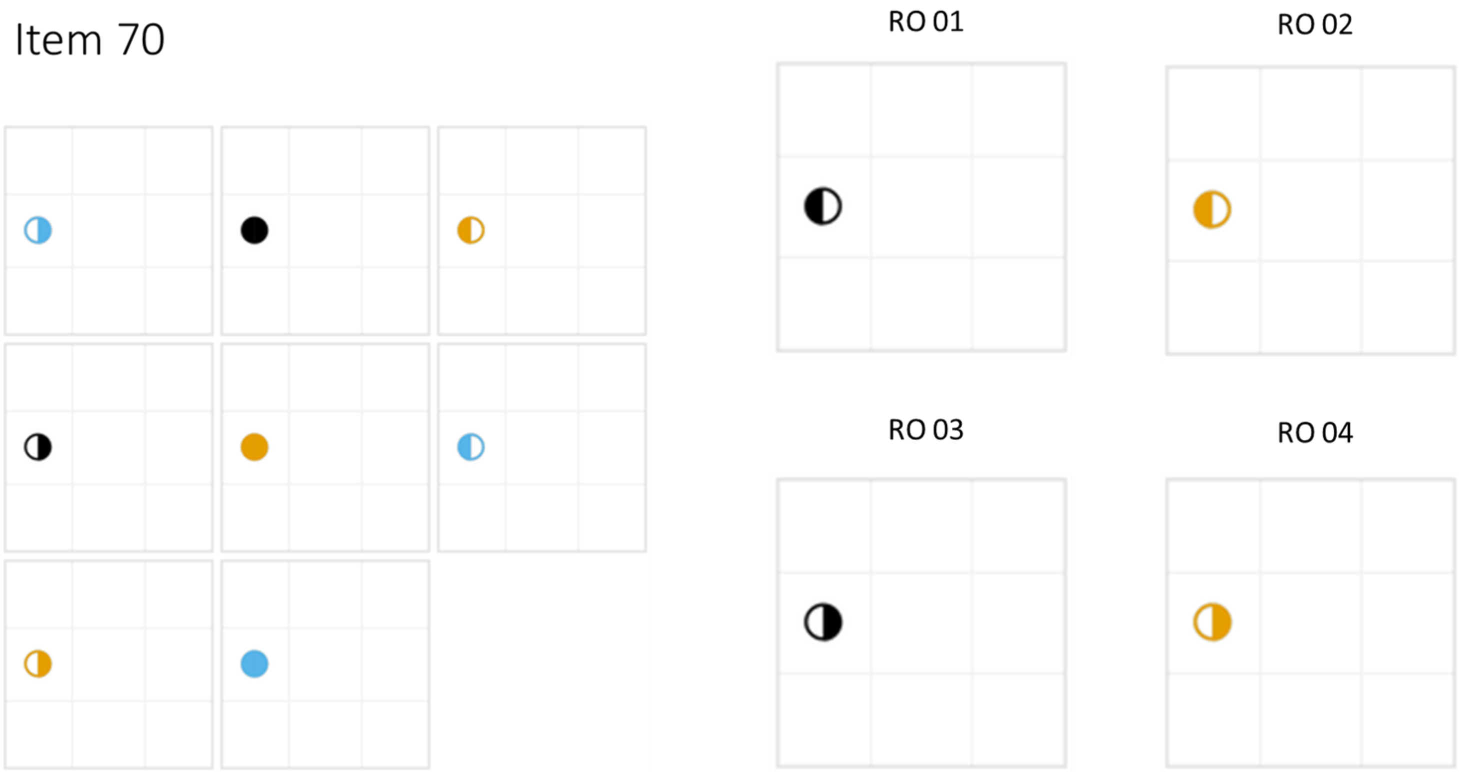
Item	β	Class Interval	θ	01	02	03	04	DK	X
CB2_70	-0.943	1	-1.998	0.276	0.034	0.207	0.034	0.345	0.103
		2	-0.064	0.697	0.028	0.220	0.025	0.015	0.015
		3	1.798	0.935	0.032	0.032	0.000	0.000	0.000

05.7 | Item 70 Empirical Distractor Curves



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh
School of Applied Psychology

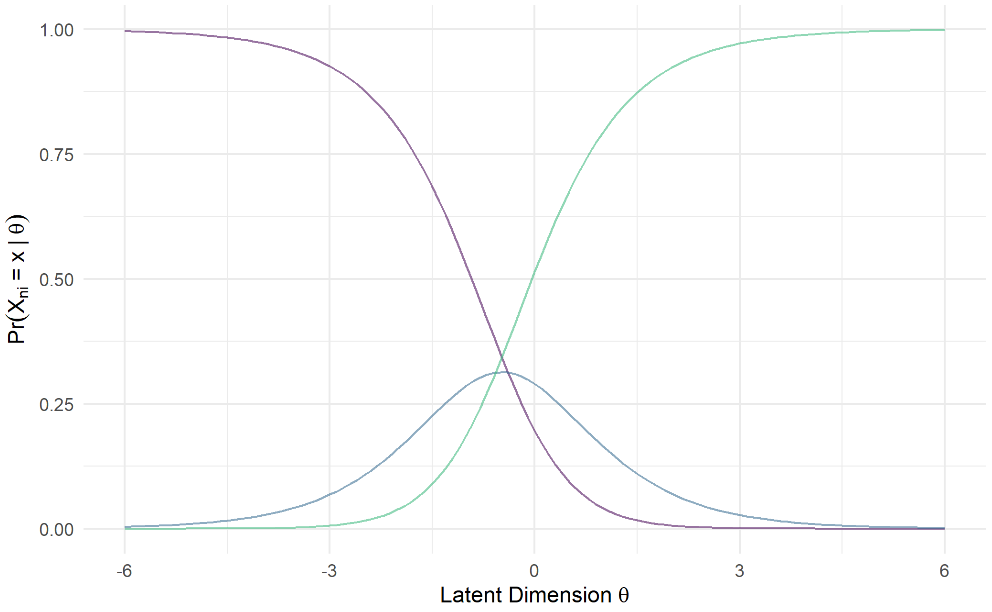
05.8 | Example Item

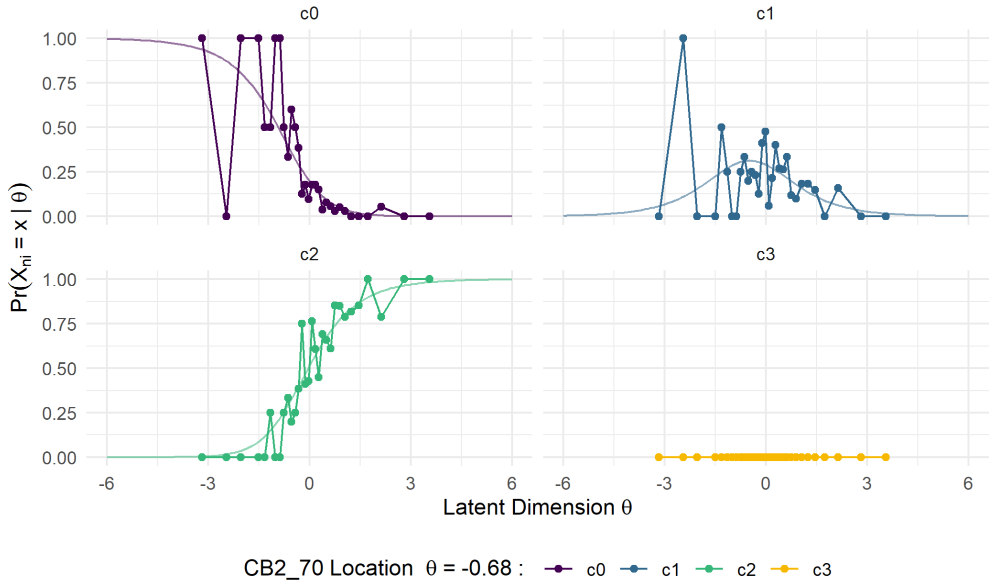


05.9 | Partial Credit Model IRF and Fit

CB2 70 Partial Credit Model Item Fit Statistics

Item	K	τ	Se	χ²	df	Infit	Outfit	Infit t	Outfit t	Discrimination
CB2_70	c1	-0.39	0.17	439.15	448	0.98	0.93	-0.18	-1	0.44
	c2	-0.96	0.16	439.15	448	0.98	0.93	-0.18	-1	0.44

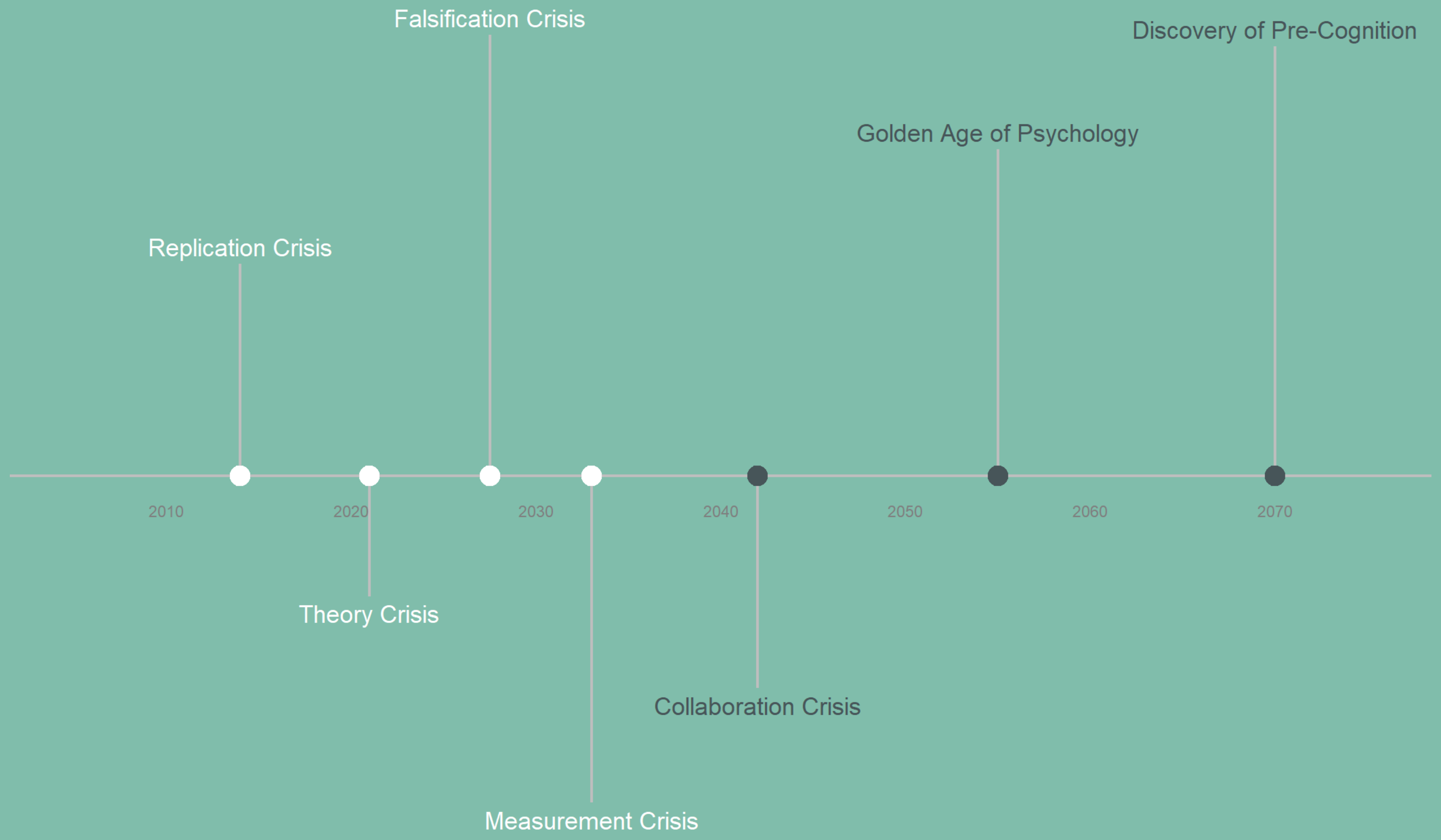




06

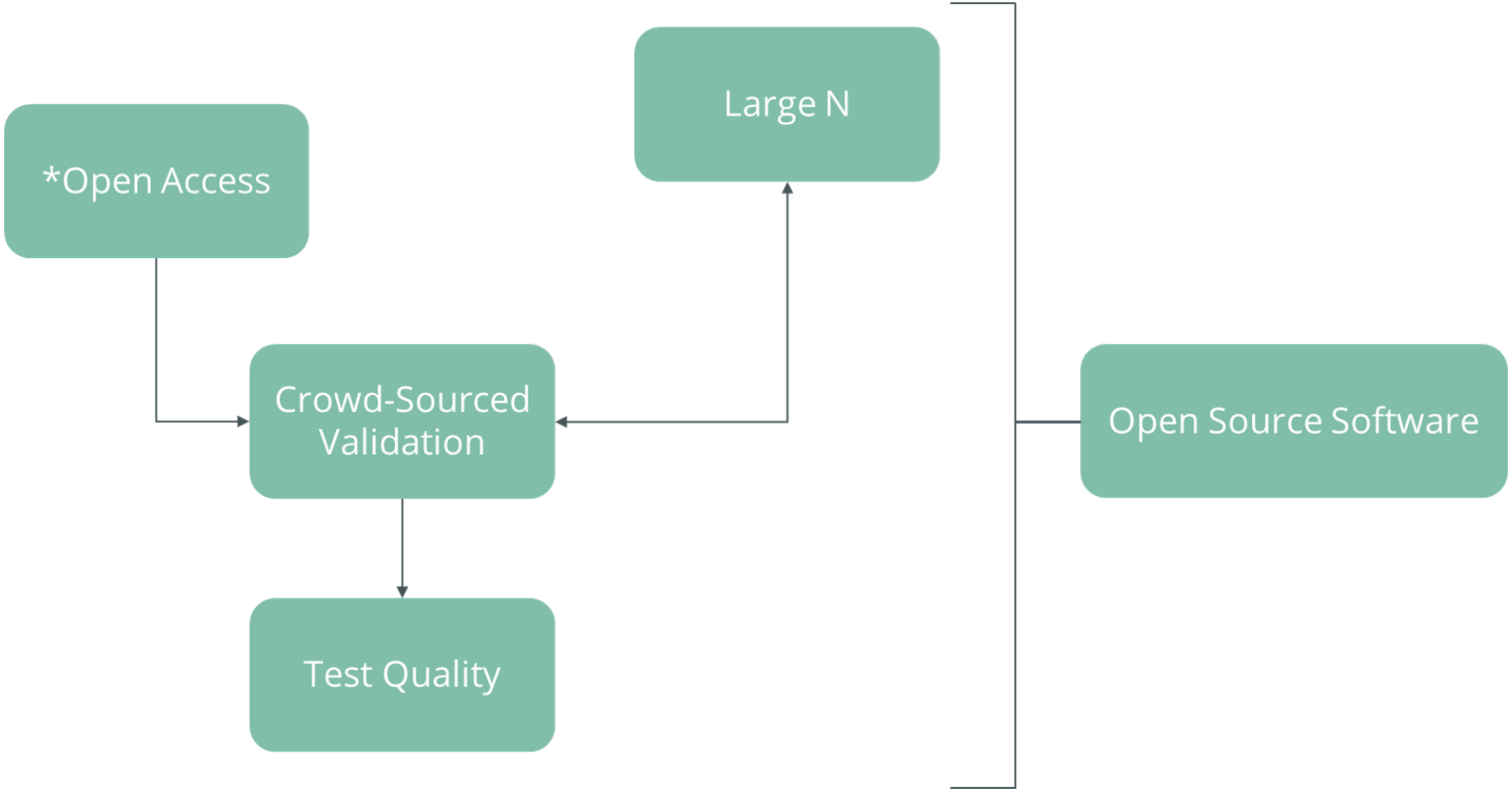
Discussion

06.1 | Crises Everywhere



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh
School of Applied Psychology

06.2 | Open Science



Open Source Software is important!

- Reproducibility
- Replicability
- Education
- Advancement of knowledge
  - Methodology
  - Theory

git

- Free-open-source-software
- Version control
- Integrated in numerous platforms

The R Project

- Free-open-source-software
- Highly developed statistical methods
- Increasing applications in psychometrics
- Included in proprietary software (SPSS)
- Highly extensible
  - Links to many other languages, e.g. Stan, C++

tidyverse

RStudio's Tidyverse

- A collection of useful R packages
- Easier to use (human readable)
- Contains powerful graphics packages (ggplot2)

07
Thank you!

Thank you!

Contact:

| SBGalvin

| shane.galvin@ucc.ie

| @S_B_Galvin

Acknowledgements:
Aaron Bolger for assistance in project development

Research supported by

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

Resources:

psketti

GitHub: https://github.com/SBGalvin/psketti
Install: `devtools::install_github("SBGalvin/psketti")`
Repository for this talk: https://github.com/SBGalvin/Graphical-Methods-for-Psychometric-Model-Selection

R
Introductory (online) text book: https://psyr.djnavarro.net/
Psychometrics in R: http://personality-project.org/r/r.guide.html

References

Andrich, D., & Styles, I (2009). Distractors with information in multiple choice items: A rationale based on the Rasch model. In E. V. Smith, Jr. & G. E. Stone (Eds.), Criterion referenced testing: Practice analysis to score reporting using Rasch measurement models (pp. 24-70). Maple Grove, MN: JAM Press.

Asril, A., & Marais, I (2011). Applying a Rasch Model Distractor Analysis: Implications for Teaching and Learning. In R.F. Cavanagh & R.F. Waugh (Eds.), Applications of Rasch Measurement in Learning Environments Research (Vol 2). (pp. 77-100). Rotterdam, Netherlands: Sense Publishers.

Borsboom, D. (2005). Measuring the mind: Conceptual issues in contemporary psychometrics. Cambridge University Press.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425

Chierchia G., Fuhrmann D., Knoll L. J., Pi-Sunyer B. P., Sakhardand, A. L., & Blakemore S. J. (2019). The matrix reasoning item bank (MaRs-IB): novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science.* 6(10), 190232.

Condon D.M., & Revelle W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*. 43, 52-64.

Flake J.K., & Fried E.I. (2019). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Preprint*: https://psyarxiv.com/hs7wm/download?format=pdf

Simon Garnier (2018). viridis: Default Color Maps from 'matplotlib'. R package version 0.5.1. https://CRAN.R-project.org/package=viridis

Hambleton, R. K., Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational measurement: issues and practice*, 12(3), 38-47.

Knoll L.J., Fuhrmann D., Sakhardande A.L., Stamp F., Speekenbrink M., Blakemore S.J. (2016). A window of opportunity for cognitive training in adolescence. *Psychological Science.*27, 1620–1631

Larner, Andrew J. (2017). Cognitive Screening Instruments: A Practical Approach. https://doi.org/10.1007/978-3-319-44775-9_1

Lakens, D. (2017). Towards a more collaborative science with StudySwap. *The 20% Statistician (BlogPost) URL http://daniellakens.blogspot.com/2017/08/*

Mair, P., Hatzinger, R., & Maier M. J. (2020). eRm: Extended Rasch Modeling. 1.0-1. https://cran.r-project.org/package=eRm

Perkel, J. M. (2020). Challenge to scientists: does your ten-year-old code still run? Nature, 584(7822), 656–658. https://doi.org/10.1038/d41586-020-02462-7

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Revelle W., Zinbarg R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. Psychometrika, 74(1), 145.

Sijtsma K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika, 74(1), 107.

Traub, R. E. (1997). Classical test theory in historical perspective. Educational Measurement: Issues and Practice, 16(4), 8–14.

Wickham, H., François, R., Henry, L., Müller, K. (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.1. https://CRAN.R-project.org/package=dplyr

Yu , CH (2020). Objective Measurement: How Rasch Modeling Can Simplify and Enhance Your Assessment. In M.S. Khine (Ed.), Rasch Measurement: Applications in Quantitative Educational Research. (pp. 47-73). Singapore Springer Singapore.